RELEVÉ DES MALADIES TRANSMISSIBLES AU CANADA

RMTC

40 ans! (1975-2015)

Le 3 septembre 2015 • VOLUME 41•9
ISSN 1719–3109

## Dans ce numéro : Données massives

Les données massives transforment peu à peu notre manière de détecter les maladies infectieuses, de les contrôler et d'en effectuer le suivi. Les données massives sont générées au moyen de l'analyse des ensembles de données volumineux, de façon à recueillir des renseignements qui n'étaient pas disponibles auparavant. Au départ, les données massives ont été lancées dans les domaines de l'analyse statistique et de l'informatique, mais elles sont dorénavant utilisées dans le cadre de la recherche scientifique. Dans le présent document, nous examinerons la façon dont les données massives découlant des ventes de produits pharmaceutiques à l'échelle nationale peuvent aider à effectuer le suivi du virus de la grippe et d'autres maladies infectieuses, la façon dont le balayage quotidien de 3 000 articles de partout dans le monde en temps presque réel peut aider à détecter les maladies infectieuses émergentes, la façon dont le séquençage du génome a amélioré notre capacité de détection pour assurer le suivi des origines des éclosions et la façon dont l'analyse de plusieurs ensembles de données peut contribuer à prévoir les futures éclosions.

---

### Surveillance

Évaluation d'un système national de surveillance syndromique en milieu pharmaceutique ..... 234
Muchaal PK, Parker S, Meganath K, Landry L, Aramini J

### Aperçu

Les données massives et le Réseau mondial d'information en santé publique (RMISP) .. 241
Dion M, AbdelMalik P, Mawudeku A

### Commentaire

Les données massives font évoluer la lutte contre les maladies infectieuses..... 248
Links M

### Actualités sur les maladies infectieuses

Les données massives et la prédiction, la prévention et le contrôle des éclosions..... 251

Données massives et éthique ..... 252

### Appel de présentations :

La soumission de manuscrits est sollicitée pour deux numéros thématiques à venir. Les dates d'échéance et sujets sont :

**Le 12 novembre 2015 – Déterminants de la santé, de l'équité en matière de santé et des maladies infectieuses**

**Le 3 décembre 2015 – Maladies infectieuses chroniques**

Vous êtes intéressés? Consultez la rubrique Renseignements à l'intention des auteurs. Si vous avez des questions, veuillez communiquer avec la rédactrice scientifique en chef à l'adresse patricia.huston@phac-aspc.gc.ca

---

Agence de la santé publique du Canada | Public Health Agency of Canada

Canada

# Relevé des maladies transmissibles au Canada

Le Relevé des maladies transmissibles au Canada (RMTC) est un journal scientifique bilingue révisé par les pairs et en accès libre en ligne publié par l'Agence de la santé publique du Canada (ASPC). Il fournit de l'information opportune et pratique sur les maladies infectieuses aux cliniciens, aux professionnels de la santé publique et aux responsables des politiques qui éclaire les politiques, le développement des programmes et les pratiques.

**Bureau de la rédaction**

Patricia Huston, M.D., M.S.P.
Rédactrice scientifique en chef

Mylène Poulin, B.SC., B.A.
Gestionnaire de la rédaction/l

Wendy Patterson
Responsable de la production
613-884-3361

Diane Staynor
Assistante à la rédaction
613-851-5033

Diane Finkle-Perazzo
Cathy Robinson
Jane Coghlan
Réviseures et correctrices d'épreuves

**Comité de rédaction du RMTC**

Michel Deilgat, C.D., B.A., M.D., CCPE
Centre des maladies infectieuses d'origine alimentaire, environnementale et zoonotique
Agence de la santé publique du Canada

Catherine Dickson, MDCM, M.Sc.
Résidente, Santé publique et médecine préventive
Université d'Ottawa

Juliana Fracassi, R.N., B.ScN.
Maladies infectieuses
Santé publique d'Ottawa

Jennifer Geduld, MHSc
Centre des maladies infectieuses d'origine alimentaire, environnementale et zoonotique
Agence de la santé publique du Canada

Judy Greig, R.N.,B.Sc., M.Sc.
Laboratoire de lutte contre les zoonoses d'origine alimentaire
Agence de la santé publique du Canada

Judy Inglis, B.Sc., MLS
Bureau du conseiller scientifique principal
Agence de la santé publique du Canada

Mohamed A. Karmali, MB ChB, FRCP(C)
Direction générale de la prévention et du contrôle des maladies infectieuses, Agence de la santé publique du Canada

Julie McGihon
Direction générale des affaires publiques et des communications
Agence de la santé publique du Canada

Robert Pless, M.D., M.Sc.
Centre de l'immunisation et des maladies respiratoires infectieuses, Agence de la santé publique du Canada

Hilary Robinson, MB ChB, M.Sc., FRCPC
Direction générale de l'infrastructure de sécurité sanitaire, Agence de la santé publique du Canada

Rob Stirling, M.D., M.Sc., MH.Sc., FRCPC
Centre de l'immunisation et des maladies respiratoires infectieuses, Agence de la santé publique du Canada

Jun Wu, Ph.D.
Centre de la lutte contre les maladies transmissibles et les infections, Agence de la santé publique du Canada

Relevé des maladies transmissibles au Canada
Agence de la santé publique du Canada
130 chemin Colonnade
Indice de l'adresse 6503B
Ottawa, (Ontario) K1A 0K9
Courriel: CCDR-RMTC@phac-aspc.gc.ca

Promouvoir et protéger la santé des Canadiens au moyen du leadership, de partenariats, de l'innovation et de la prise de mesures dans le domaine de la santé publique.
Agence de la santé publique

Publication autorisée par le ministre de la Santé.
© Sa Majesté la Reine du Chef du Canada, représentée par le ministre de la Santé, 2015
ISSN 1719-3109
Pub 150006

On peut aussi consulter cette publication en ligne: http://www.phac-aspc.gc.ca/publicat/ccdr-rmtc/15vol41/index-fra.php
Also available in English under the title: *Canada Communicable Disease Report*

Surveillance

# Évaluation d'un système national de surveillance syndromique en milieu pharmaceutique

**Muchaal PK[1]*, Parker S[1], Meganath K[1], Landry L[1], Aramini J[2]**

[1]Centre des maladies infectieuses d'origine alimentaire, environnementale et zoonotique, Agence de la santé publique du Canada, Guelph (Ontario)
[2]Intelligent Health Solutions, Fergus (Ontario)

*Correspondance : pia.muchaal@phac-aspc.gc.ca

## Résumé

**Contexte :** La surveillance de la santé publique traditionnelle fournit de l'information précise mais généralement pas en temps opportun. De nouveaux systèmes d'alerte rapide qui exploitent des données électroniques en temps opportun sont mis en œuvre, mais la valeur pour la santé publique de ces systèmes demeure en grande partie méconnue.

**Objectif :** Évaluer la rapidité et l'exactitude des données sur les ventes de produits pharmaceutiques pour les infections respiratoires et gastro-intestinales et déterminer leur utilité dans le cadre du soutien de la surveillance des maladies gastro-intestinales.

**Méthodologie :** Pour évaluer la rapidité, une analyse prospective et rétrospective du flux de données a été utilisée pour comparer les caractéristiques chronologiques de chaque suite de données. Aux fins d'évaluation de l'exactitude, les ordonnances d'antiviraux en Ontario ont été comparées à celles de cas confirmés de grippe et de syndrome grippal (SG) du mois d'août 2009 au mois de janvier 2015. De plus, les ventes de produits en vente libre pour les maladies respiratoires de la Nouvelle-Écosse ont été comparées aux rapports de laboratoire concernant les détections d'agents pathogènes des voies respiratoires du mois de janvier 2014 au mois de mars 2015. Les données sur les éclosions de maladies entériques (2011-2014) de la Nouvelle-Écosse ont été comparées à celles des ventes de produits gastro-intestinaux pendant la même période. Pour évaluer l'utilité, les ventes de produits pharmaceutiques gastro-intestinaux ont été surveillées à l'échelle du Canada pour détecter des augmentations inhabituelles, des rapports ont été acheminés aux provinces et aux territoires une fois par semaine du mois de décembre 2014 au mois de mars 2015 et une évaluation de suivi auprès des intervenants a été effectuée.

**Résultats :** Le nombre d'ordonnances d'antiviraux en Ontario rédigées entre 2009 et 2015 correspondait étroitement aux dates d'apparition de la maladie et à l'ampleur des cas de grippe confirmés. Les ventes de produits en vente libre pour les maladies respiratoires en Nouvelle-Écosse correspondaient aux augmentations des agents pathogènes des voies respiratoires non liés à la grippe au sein de la collectivité. Aucune corrélation définitive n'a été cernée entre les éclosions de maladies entériques et les ventes de médicaments gastro-intestinaux en vente libre en Nouvelle-Écosse. L'évaluation de la surveillance à l'échelle nationale a révélé une augmentation significative des ventes de produits gastro-intestinaux qui pourrait être liée aux éclosions dans plus d'une province ou d'un territoire.

**Conclusion :** La surveillance des ordonnances de médicaments en milieu pharmaceutique et des ventes de médicaments en vente libre peut fournir un complément précis et ponctuel aux activités de surveillance traditionnelle de la santé publique liée aux maladies respiratoires, mais l'évaluation initiale n'a pas indiqué que le suivi des ventes de médicaments gastro-intestinaux en vente libre a été utile pour repérer des éclosions de maladies entériques dans plus d'une province ou d'un territoire au cours de la période de l'étude.

## Introduction

Au Canada, la surveillance traditionnelle de la santé publique liée aux maladies infectieuses repose largement sur la déclaration des cas confirmés en laboratoire. Ce mécanisme fournit des renseignements fiables, mais un délai considérable peut s'écouler entre le dépistage et la déclaration de la maladie aux autorités de santé publique. De ce fait, l'occasion de mettre en place des interventions est perdue. En outre, pour les maladies infectieuses qui sont généralement associées à une maladie variant de bénigne à modérée, le traitement est souvent empirique (aucun test n'est effectué), ce qui limite la capacité de surveillance en laboratoire afin de fournir une évaluation précise de la maladie au sein de la collectivité.

La surveillance syndromique comprend souvent les données volumineuses; elle repose sur l'utilisation d'indicateurs de la santé ou de mesures de substitution non spécifiques (p. ex. absentéisme scolaire, ventes de médicaments, appels au service télésanté) afin de fournir un diagnostic provisoire (ou « syndrome »). Ces sources de données ont tendance à être non spécifiques et sensibles et rapides et elles peuvent augmenter et complémenter les renseignements fournis par les systèmes traditionnels de surveillance des diagnostics en fonction des tests (1).

Au cours des dix dernières années, la surveillance en milieu pharmaceutique est devenue une nouvelle capacité en matière de santé publique au Canada et à l'échelle internationale (2, 3,4,5). L'Agence de la santé publique du Canada (l'Agence) a mis en œuvre une surveillance syndromique en milieu pharmaceutique pour la première fois en 2004, de façon à évaluer la faisabilité de la surveillance des produits gastro-intestinaux en vente libre comme outil de détection précoce des maladies ou des éclosions communautaires d'origine alimentaire et hydrique (2). En 2009, l'Agence a de nouveau fait appel à la surveillance syndromique en milieu pharmaceutique en réponse à l'éclosion de grippe pandémique pH1N1. Une analyse rétrospective de la deuxième vague de grippe pH1N1 a démontré que la surveillance en milieu pharmaceutique offrait un mécanisme efficace pour surveiller et détecter l'activité de type influenza plus rapide que les systèmes traditionnels de surveillance (6).

À la suite de la pandémie de grippe H1N1, le système de surveillance en milieu pharmaceutique a été étendu pour déterminer l'utilité durant la mise en œuvre du *plan stratégique de surveillance de l'Agence*. Un certain nombre de médicaments ont fait l'objet d'un suivi, y compris des analgésiques, les médicaments contre les allergies et les ordonnances de médicaments antidépresseurs et de médicaments pour les maladies cardiovasculaires.

L'objectif de cette étude était d'évaluer trois aspects de ce système, soit l'exactitude, la rapidité et l'utilité de la surveillance en milieu pharmaceutique pour appuyer la surveillance de la grippe saisonnière et des maladies respiratoires et la détection des éclosions multijuridictionnelles de maladies entériques (c.-à-d. des maladies touchant plus d'une province ou d'un territoire).

## Méthodologie

### Sources des données

Des données ont été obtenues au sujet de la surveillance des ordonnances d'antiviraux liées aux maladies respiratoires en Ontario (2009-2015) et des ventes de médicaments contre les maladies respiratoires en vente libre en Nouvelle-Écosse (2014-2015). Elles ont été comparées à des cas confirmés de grippe et à des rapports de syndrome grippal en Ontario et à des rapports de détection en laboratoire des agents pathogènes des voies respiratoires non liés à la grippe en Nouvelle-Écosse.

Des données ont été obtenues au sujet de la surveillance des ventes de produits gastro-intestinaux en vente libre pour les maladies gastro-intestinales aiguës dans l'ensemble du Canada et les données provenant de la Nouvelle-Écosse ont été comparées aux données concernant l'éclosion locale déclarées par les autorités provinciales de la santé.

### Collecte et soumission des données

L'acquisition de données pharmaceutiques a été établie par l'entremise d'un contrat conclu avec un partenaire de l'industrie, soit Rx Canada. Les données quotidiennes sur les ventes de médicaments sur ordonnance proviennent de 13 chaînes de pharmacies à l'échelle nationale et quatre pharmacies indépendantes représentant plus de 3 000 magasins à l'échelle du pays. Des données sur les produits en vente libre ont été recueillies chaque jour par les détaillants de six chaînes de pharmacies représentant 1 863 magasins dans l'ensemble du Canada (à l'exception du Nunavut). Les ensembles de données sur les médicaments en vente libre et les médicaments sur ordonnance couvrent plus de 85 % des régions sanitaires au Canada.

Les produits pharmaceutiques étaient classés en fonction des syndromes en regroupant les produits dans la catégorie des maladies respiratoires ou la catégorie des maladies gastro-intestinales. Les données sur les ventes de produits en vente libre ont été normalisées en magasin avant le regroupement en fonction des services de santé publique, de la province ou du territoire et du pays, afin de réduire au minimum les effets du délai et de la fréquence de transmission entre les magasins et les chaînes de vente au détail. La proportion de ventes quotidiennes de certains produits pour guérir une maladie entérique aiguë a été calculée et présentée en fonction des ventes de tous les autres produits en vente libre. La moyenne mobile de sept jours normalisée des sous-catégories de maladies gastro-intestinales divisée par les données des autres médicaments en vente libre est illustrée par un graphique sur une échelle logarithmique. Un simple algorithme d'avertissement reposant sur les écarts-types 1, 2 et 3 supérieurs à la moyenne mobile de sept jours a été utilisé pour détecter les augmentations aberrantes de ventes de produits pharmaceutiques.

### Réponse en temps opportun

L'analyse prospective et rétrospective de différents flux de données mis à profit dans le cadre de cette étude (produits en vente libre, ordonnances, syndrome grippal, déclaration en laboratoire) a été comparée pour évaluer la rapidité de la surveillance de la santé publique.

### Exactitude

L'exactitude de la pharmacosurveillance pour la grippe et les syndromes grippaux a été évaluée par l'entremise d'une analyse rétrospective comparant les données des produits pharmaceutiques et les données de surveillance. Les données de surveillance ont été extraites à partir du système Surveillance de l'influenza. Les données des ventes liées aux ordonnances d'antiviraux sur cinq ans en Ontario ont été comparées à celles des cas déclarés de syndrome grippal (SG) entre le mois d'août 2009 et le mois de janvier 2015 et à celles des cas confirmés de grippe entre les années 2011 et 2015. De concert avec l'analyse descriptive, le coefficient de corrélation de Spearman (*rho*) a été déterminé pour les trois périodes (de janvier à décembre, de novembre à mars et d'avril à octobre).

Afin d'évaluer l'exactitude du suivi d'autres infections des voies respiratoires, les détections hebdomadaires en laboratoire à l'échelle provinciale des pathogènes viraux des voies respiratoires provenant de la Nouvelle-Écosse signalées entre le mois de janvier 2014 et le mois de mars 2015 ont été évaluées par rapport aux ventes de médicaments en vente libre liés aux maladies respiratoires pour la même période. Les données sur les médicaments en vente libre liés aux maladies respiratoires ont été regroupées par semaine et des variables de semaine ayant un délai d'une à cinq semaines ont été créées. Les corrélations utilisant les coefficients de Pearson ont été calculées entre les ventes de médicaments en vente libre liés aux maladies respiratoires, les détections d'agents viraux des voies respiratoires non liés à la grippe et les tests pour le virus respiratoire syncytial (VRS).

Pour évaluer l'exactitude de la pharmacosurveillance des maladies entériques, les ventes de médicaments gastro-intestinaux en vente libre ont été comparées aux données sur les éclosions de maladies entériques en Nouvelle-Écosse pendant la période de 2011 à 2014, à l'aide d'une analyse descriptive.

### Utilité

Les rapports de surveillance pharmaceutique fondés sur les ventes de produits gastro-intestinaux en pharmacie partout au Canada ont été générés et diffusés chaque semaine aux intervenants provinciaux et

territoriaux. Des renseignements et des tendances régionales en matière de santé publique à l'échelle provinciale se trouvaient dans ces rapports. L'utilité de l'information figurant dans les rapports à l'intention des intervenants a été évaluée à l'aide d'un sondage élaboré à l'aide de FluidSurveys.

Les logiciels SAS 9.3 et Stata 13 ont été utilisés pour l'exécution des examens analytiques.

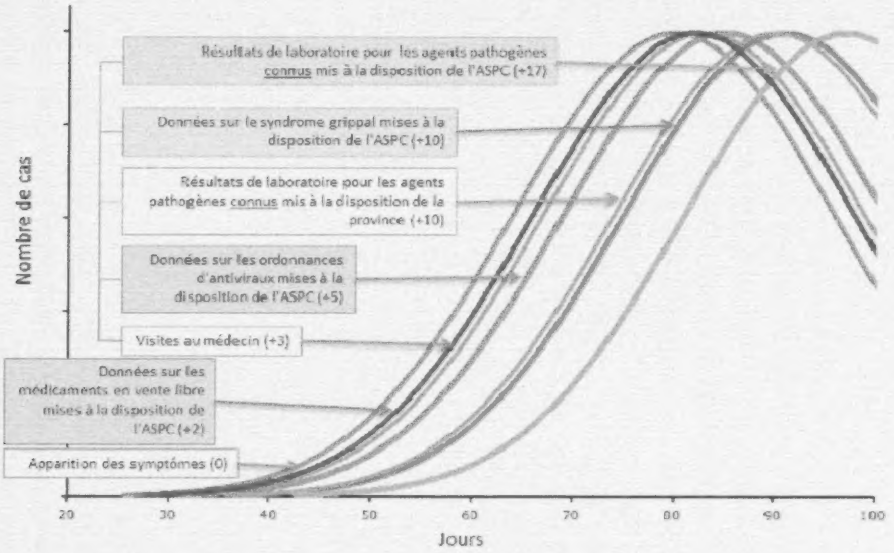
## Résultats

### Réponse en temps opportun

Les données sur les ventes de produits pharmaceutiques (p. ex. les médicaments sur ordonnance et les produits en vente libre) étaient disponibles en temps quasi réel, soit environ 48 heures après une transaction. L'Agence avait accès aux données sur les ventes de produits pharmaceutiques environ cinq à huit jours plus tôt que les rapports de médecin pour un syndrome grippal, de 10 à 12 jours avant les confirmations de laboratoire de la grippe et jusqu'à 17 jours avant l'émission des rapports de détections des agents viraux des voies respiratoires.

La **Figure 1** illustre les délais quant aux données pharmaceutiques, cliniques et de laboratoire par rapport à l'estimation de l'apparition de la maladie et à la disponibilité de l'information aux fins de surveillance des maladies respiratoires.

**Figure 1 : Rapidité de la surveillance des maladies respiratoires graves**



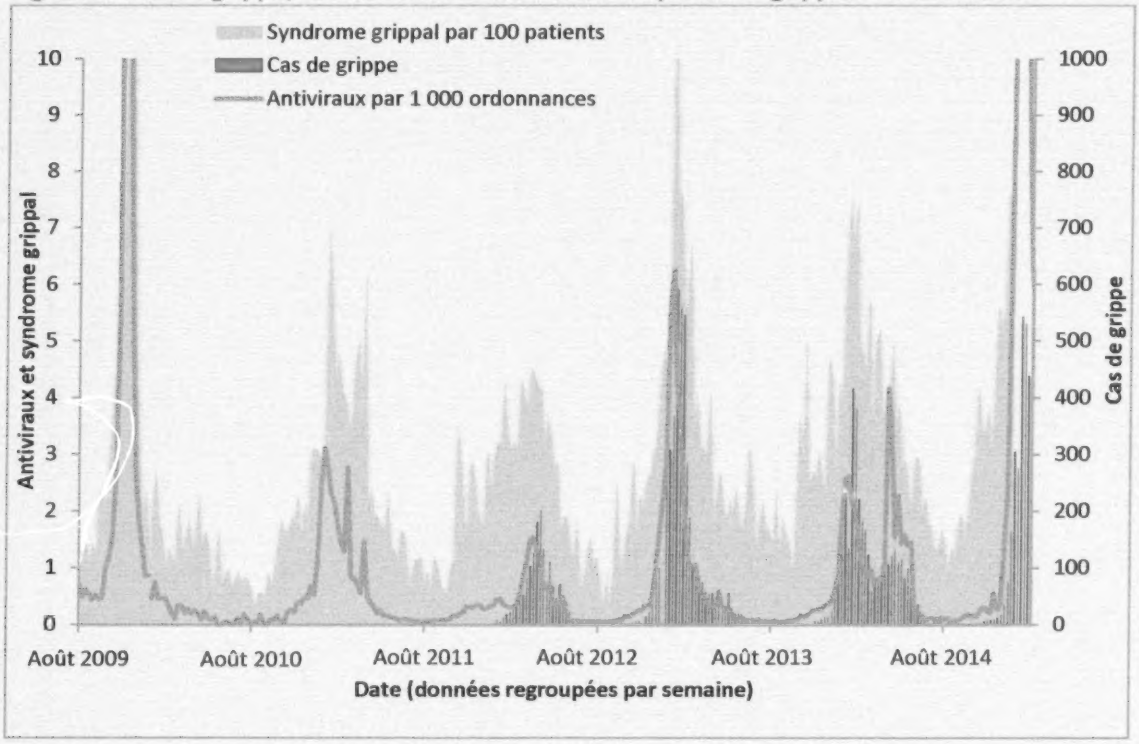

ASPC : Agence de la santé publique du Canada

### Exactitude

*Surveillance des maladies respiratoires :* Chaque semaine, les ordonnances d'antiviraux délivrées des médicaments oseltamivir et zanamivir (proportion de toutes les ordonnances délivrées) ont été comparées au taux de syndrome grippal (cas de syndrome grippal et nombre total de visites des patients) et au nombre de cas de grippe confirmés en laboratoire en Ontario (**Figure 1**). Les ventes d'antiviraux entre 2009 et 2015 correspondaient étroitement aux dates d'apparition de la maladie et aux cas de grippe confirmés. Le *rho* de Spearman était de 0,90, 0,93 et 0,83 (toutes les valeurs étaient de $p < 0,001$) pour la période de janvier à décembre, de novembre à mars et d'avril à octobre, respectivement. De plus, l'ampleur des ventes d'antiviraux était parallèle au fardeau des cas confirmés. Le coefficient de corrélation entre les cas confirmés

et le syndrome grippal n'était pas aussi fort. Pour janvier et décembre, le *rho* était de 0,80, tandis qu'il était de 0,70 en novembre et en mars et de 0,60 d'avril à octobre. Les rapports de syndrome grippal ont suivi la même tendance que les cas confirmés et les antiviraux (**Figure 2**). Toutefois, les taux de syndrome grippal variaient considérablement à la fin du printemps et en été.

**Figure 2 : Cas de grippe, ordonnances d'antiviraux et syndrome grippal en Ontario**




Les résultats de l'analyse de la Nouvelle-Écosse ont révélé des ventes de médicaments respiratoires en vente libre associées à des augmentations des pathogènes des voies respiratoires non liés à la grippe dans la collectivité (virus respiratoire syncytial, rhinovirus, coronavirus, parainfluenza, adénovirus, métapneumovirus humain). Les ventes de médicaments en vente libre correspondaient au nombre d'analyses du virus respiratoire syncytial quatre semaines plus tard ($r = 0,76$, $p < 0,001$). Les ventes de médicaments en vente libre correspondaient au nombre de détections d'autres agents viraux des voies respiratoires deux semaines plus tard (0,74, $p < 0,001$) (**Figure 2**).

*Surveillance des maladies gastro-intestinales* : 262 éclosions ont été déclarées en Nouvelle-Écosse entre 2012 et 2014. De ce nombre, 66 % des éclosions étaient imputables au norovirus, 82 % étaient principalement causées par la transmission interhumaine, et la majorité (81 %) est survenue dans des établissements résidentiels. Aucune corrélation définitive n'a été détectée entre les éclosions et les ventes de médicaments gastro-intestinaux en vente libre.

## Utilité

Au cours de la phase pilote, aucune augmentation importante des ventes de produits gastro-intestinaux pouvant être liée de façon certaine à une éclosion multijuridictionnelle de maladies entériques n'a été détectée. Bien que seulement 40 % des intervenants aient répondu à l'enquête d'évaluation, la plupart (87 %) d'entre eux ont indiqué que les renseignements pharmaceutiques étaient importants pour leur territoire de compétence. Cinq territoires de compétence ont indiqué qu'ils avaient utilisé ces

renseignements. Parmi ceux-ci, 20 % les ont utilisés pour lancer une enquête, 75 % les ont utilisés pour détecter un événement de santé, 20 % ont transmis l'information aux cliniciens, et ils les ont tous utilisés pour connaître la situation.

## Discussion

La pharmacosurveillance nationale et provinciale semblait être plus utile pour la détection précoce des maladies respiratoires que pour la détection provinciale des maladies entériques. Les ordonnances d'antiviraux étaient une marque claire de l'activité grippale pour la surveillance des maladies respiratoires. La valeur des médicaments en vente libre pour les maladies respiratoires était moins claire, mais ils pourraient être utiles pour la surveillance d'autres virus respiratoires. Bien que les ventes de produits gastro-intestinaux représentent une bonne marque pour les infections saisonnières au norovirus dans la collectivité (6), la valeur de la pharmacosurveillance pour l'activité entérique liée à une éclosion était moins certaine.

Les points forts de cette étude comprennent la représentativité nationale des données et la documentation précise sur la rapidité d'obtention des données de pharmacosurveillance. Un point faible possible de l'étude est qu'elle incluait seulement les médicaments prescrits et les médicaments en vente libre des pharmacies de détail non liées à des établissements de soins de santé. Par exemple, bien que les données sur les éclosions de maladies entériques en Nouvelle-Écosse soient solides, la majorité des éclosions de norovirus prises en compte durant la période de l'étude ont eu lieu dans des établissements de soins de longue durée. Les médicaments offerts dans les établissements de soins de longue durée peuvent provenir de pharmacies ne contribuant pas de données sur les ventes au projet-pilote. De plus, les achats en vrac de médicaments en vente libre par les établissements n'étaient peut-être pas pris en compte dans les données fournies.

Le transfert automatique des données du point de vente appuie l'acquisition en temps réel ou quasi réel de données sur la santé propres à une collectivité. Les données des pharmacies sur les ordonnances et les médicaments en vente libre fournissent une plus grande rapidité d'obtention des données et une plus grande résolution géographique, ainsi qu'une meilleure couverture nationale, que les autres données sur la santé actuellement utilisées. Il a été démontré que les données sur les ventes de produits pharmaceutiques appuient les données sur la grippe saisonnière et contribuent à la prise de décisions concernant l'allocation des ressources. Les données sur les ordonnances et les médicaments en vente libre permettent aussi une intervention précoce par rapport aux sources de données traditionnelles seulement. De plus, les données pharmaceutiques peuvent être utilisées pour appuyer la surveillance et l'intervention à plusieurs fins de santé publique (p. ex. santé mentale, tendances des ordonnances des médecins et maladies chroniques) et par de nombreux organismes et différents territoires de compétence. Comme c'est le cas pour les autres systèmes de surveillance, des investissements dans les ressources technologiques et analytiques nécessaires doivent être faits pour mener une surveillance syndromique continue en milieu pharmaceutique.

## Conclusion

Une pharmacosurveillance exacte et en temps opportun peut améliorer la capacité de la santé publique à détecter et à quantifier l'activité à l'échelle locale, multijuridictionnelle et nationale. D'autres recherches sont nécessaires pour déterminer dans quelles conditions elle est le plus utile et pour la comparer à d'autres stratégies de surveillance en temps réel.

## Remerciements

Nous souhaitons remercier les provinces et les territoires d'avoir fourni les données de cas pour l'analyse et d'avoir participé à l'étude pilote, de même que les responsables du rapport de *Surveillance de l'influenza* d'avoir fourni les données sur la surveillance des maladies respiratoires. L'important engagement et le soutien des détaillants de chaînes de pharmacies, des pharmacies et de Rx Canada ont joué un rôle essentiel dans la réussite des diverses phases du projet-pilote de surveillance syndromique en milieu pharmaceutique.

## Conflit d'intérêts

Jeffery Aramini est un employé d'Intelligent Health Solutions, l'entreprise qui a fourni l'analyse des données.

## Financement

Le soutien et le financement du projet de pharmacosurveillance ont été fournis par l'Agence de la santé publique du Canada.

## Références

(1) Berger M, Shiau R, Weintraub JM. Review of syndromic surveillance: Implications for waterborne disease detection. J Epidemiol Community Health. 2006 Jun;60(6):543–550.

(2) Chadwick D. The Rhode Island Department of Health. The first statewide system for tracking disease using prescription data. Press Release Archives. State of Rhode Island. Department of Health. 2009. Available from: http://www.ri.gov/press/view/10017.

(3) Pavlin JA, Murdock P, Elbert E. Conducting population behavioral health surveillance by using automated diagnostic and pharmacy data systems. MMWR. 2004;53:166–172.

(4) Sugawara T, Ohkusa Y, Ibuka Y, Kawanohara H, Taniguchi K, et al. Real-time prescription surveillance and its application to monitoring seasonal influenza activity in Japan. J Med Internet Res. 2012;14(1):e14. Available from: http://www.jmir.org/2012/1/e14/.

(5) Van den Wijngaard C, van Pelt W, Nagelkerke N, Kretzschmar M, Koopmans M. Evaluation of syndromic surveillance in the Netherlands: Its added value and recommendations for implementation. Euro Surveill. 2011;16(9):19806. Available from http://www.eurosurveillance.org/images/dynamic/EE/V16N09/art19806.pdf

(6) Aramini J, Muchaal PK, Pollari F. Value of pharmacy-based influenza surveillance — Ontario, Canada, 2009. MMWR. 2013 May 24; 62(20);401-404.

Aperçu

# Les données massives et le Réseau mondial d'information en santé publique (RMISP)

**Dion M[1], AbdelMalik P[2], Mawudeku A[2*]**

[1] Schulich School of Family Medicine and Dentistry, Université de Western Ontario, London (Ontario)
[2] Centre de mesures et d'interventions d'urgence, Agence de la santé publique du Canada, Ottawa (Ontario)

* Correspondance : Abla.Mawudeku@phac-aspc.gc.ca

## Résumé

**Contexte :** La mondialisation et le risque de propagation rapide des maladies infectieuses émergentes ont rendu de plus en plus nécessaires la surveillance continue et la détection précoce de ces maladies. Le Réseau mondial d'information en santé publique (RMISP) a été établi pour améliorer la connaissance de la situation et accroître la capacité de détection précoce des événements de santé publique émergents.

**Objectif :** Décrire comment le RMISP utilise les données massives comme technique efficace de détection précoce des éclosions de maladies infectieuses et déterminer les orientations futures possibles pour le RMISP.

**Constatations :** Chaque jour, le RMISP analyse plus de 20 000 nouvelles en ligne (provenant de plus de 30 000 sources) en neuf langues dans le monde entier. Un programme sur Internet compile les données à partir d'un algorithme qui repère les signaux potentiels d'événements de santé publique émergents, lesquels sont ensuite examinés par une équipe multilingue et pluridisciplinaire. Une alerte est envoyée si un risque potentiel est détecté. Ce processus a fait ses preuves durant l'éclosion du syndrome respiratoire aigu sévère (SRAS) et a été adopté peu après par un certain nombre de pays afin de satisfaire au nouveau Règlement sanitaire international (RSI), qui exige que chaque pays soit capable de détecter et de signaler de façon précoce les maladies. Le RMISP a repéré l'éclosion précoce de SRAS en Chine, s'est vu attribuer le mérite de la première alerte du coronavirus du syndrome respiratoire du Moyen-Orient (CoV-SRMO), et a joué un rôle important dans le suivi de l'éclosion du virus Ebola en Afrique occidentale. De futures évolutions sont envisagées pour faire progresser la capacité du RMISP à tirer parti d'autres sources de données massives, telles que les médias sociaux, ainsi que sa capacité analytique en matière d'élaboration d'algorithmes.

**Conclusion :** L'adoption précoce des données massives par le RMISP a permis d'accroître la capacité mondiale à détecter les éclosions internationales de maladies infectieuses et les autres événements de santé publique. L'intégration de sources supplémentaires de données massives et les progrès en matière de capacité analytique pourraient renforcer encore davantage la capacité du RMISP à détecter les maladies et à lancer l'alerte rapidement.

## Introduction

Avec l'avancée de la mondialisation, les maladies transmissibles et les événements de santé publique émergents se propagent de plus en plus rapidement. Par conséquent, leur surveillance continue et leur détection précoce sont encore plus importantes pour prévenir ou atténuer la propagation internationale des maladies infectieuses et pour laisser aux pays le temps nécessaire pour s'y préparer et y faire face. Les données massives désignent des ensembles de données extrêmement volumineux provenant de sources telles que les médias sociaux ou les journaux, qui nécessitent des méthodes de calcul puissantes pour révéler des tendances, des schémas ou des prédictions de probabilité d'un événement (1,2). Les données massives sont utilisées pour optimiser les processus de vente et d'affaires, orienter les échanges de joueurs entre les équipes sportives, ou encore améliorer l'urbanisme. Elles deviennent rapidement essentielles à

divers aspects de la santé, allant de l'administration des soins de santé à l'application Suivi de la grippe de Google et à la pharmacosurveillance (3).

Le Canada a été l'un des premiers à adopter les données massives pour le repérage initial des infections émergentes, et ce, dès 1997 avec la mise en place du Réseau mondial d'information en santé publique (RMISP), fruit d'un effort de coopération entre (à l'époque) Santé Canada et l'Organisation mondiale de la Santé (OMS) (4,5). Le RMISP continue à être géré par l'Agence de la santé publique du Canada (l'Agence) et rassemble un réseau mondial de professionnels de la santé publique et d'organisations (p. ex. ministres de la Santé) dans une optique de connaissance de la situation et de détection précoce des événements de santé publique émergents. Le RMISP s'appuie sur un système automatique sur Internet qui parcourt les journaux et les autres communications dans le monde entier pour repérer les indicateurs potentiels (ou « signaux ») d'éclosions, qui sont ensuite analysés et rapidement évalués par une équipe multilingue et pluridisciplinaire à l'Agence. Lorsqu'un risque est détecté, les analystes diffusent les renseignements pertinents et alertent les hauts fonctionnaires et les intervenants pour que ceux-ci prennent une décision. Bien qu'il ait été initialement conçu pour repérer les éclosions de maladies transmissibles, le système est également utilisé pour surveiller les dangers chimiques et radionucléaires potentiels (4,6).

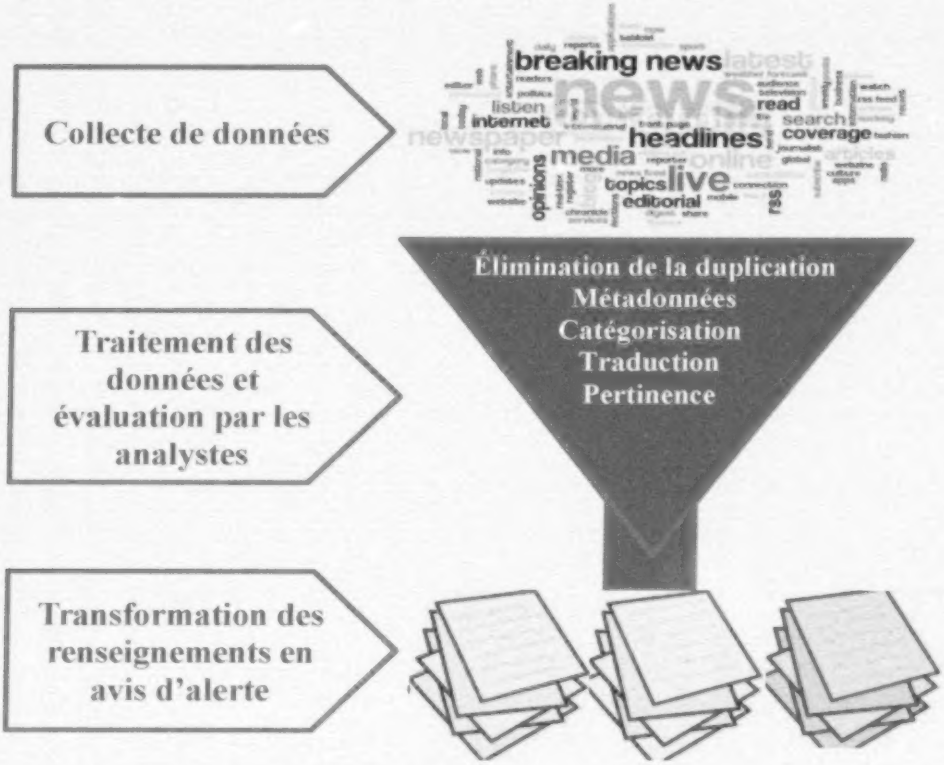
Le présent article a pour objectif de décrire le fonctionnement du RMISP dans le contexte des données massives, de l'illustrer par des exemples récents du RMISP en action et d'explorer les possibles orientations futures.

## Le RMISP et les données massives

Les données massives sont définies par trois V : volume, vitesse et variété (7,8). Le volume décrit la quantité de données recueillies, la vitesse correspond à la fréquence à laquelle les données sont recueillies et diffusées, et la variété désigne la multiplicité des sources utilisées pour compiler les données (7).

Le volume et la variété du RMISP sont illustrés par l'utilisation de fonctions de recherche et d'agrégateurs de nouvelles (sociétés donnant accès à des milliers de sources d'actualité dont le contenu est automatiquement indexé) qui rassemblent de grandes quantités d'ensembles de données provenant de multiples sources différentes. Une application sur Internet du système du RMISP balaye et analyse en continu les sources de nouvelles acquises en neuf langues (arabe, anglais, persan, français, portugais, russe, chinois simplifié, espagnol et chinois traditionnel) (4). La quantité de données générées dépend des critères, des variables et des algorithmes définis pour les agrégateurs (6). Ces algorithmes repèrent les signaux potentiels d'événements de santé publique émergents et filtrent les données non pertinentes considérées comme du « bruit » (**Figure 1**) (7). Chaque jour, en moyenne, le RMISP traite 3 000 nouvelles (9). Le volume augmente lorsque les sources d'actualité élargissent leur couverture des événements de santé publique émergents, comme dans le cas de la récente éclosion du virus Ebola en Afrique occidentale.

**Figure 1 : Le flux d'information dans le processus du RMISP**




Le RMISP dispose d'une abondante variété de sources de données. Les agrégateurs de nouvelles du RMISP s'appuient sur une grande diversité de journaux nationaux et locaux et de bulletins d'information provenant du monde entier (4,6). Les journaux et bulletins d'information locaux sont balayés, car les événements émergents peuvent être un phénomène localisé et sont souvent mentionnés dans les publications communautaires. Diverses rubriques des publications d'actualité (sports, voyages et finances) sont également surveillées, car elles peuvent signaler un événement de santé publique émergent. Une lecture en plusieurs langues est effectuée afin de repérer les événements de santé publique qui ne sont pas signalés dans les nouvelles en anglais (10).

Après application des algorithmes au sein du système du RMISP, environ 60 % des données (1 800 nouvelles) sont considérées comme des événements de santé publique pertinents aux fins d'évaluation. Les analystes du RMISP parcourent ces actualités pour repérer les événements pouvant avoir des répercussions et émettre des alertes aux fins de prise de décision par les intervenants. Ont accès au système du RMISP les organismes qui ont la responsabilité de surveiller, contrer et/ou atténuer les menaces de santé publique émergentes. Le RMISP réunit des ministères de la santé, d'autres ministères et organismes gouvernementaux, des organisations internationales et non gouvernementales, ainsi que des sociétés privées.

La capacité du RMISP sur le plan de la vitesse est impressionnante. Il fonctionne en temps quasi réel, 24 heures sur 24 et 7 jours sur 7 (4). Le système du RMISP récupère des données pertinentes auprès des agrégateurs de nouvelles toutes les 15 minutes et est capable d'achever le traitement (y compris la traduction) des données en moins d'une minute (9).

## Le RMISP en action

### Détection précoce

Le RMISP s'est avéré une ressource efficace pour la détection précoce des éclosions de maladies infectieuses. Son utilité a été initialement démontrée pendant l'éclosion du syndrome respiratoire aigu sévère (SRAS) en 2003, lorsque les premières alertes ont été lancées dans les actualités des journaux chinois. La première nouvelle en anglais concernant une éclosion atypique en Chine a été mentionnée par une société pharmaceutique dans la rubrique financière d'un journal qui avait fait état d'une hausse des ventes de médicaments antiviraux (11). Cela a permis non seulement de signaler l'émergence de l'éclosion, mais également de fournir des renseignements supplémentaires sur l'utilisation locale des antiviraux pour contenir la propagation du virus.

À la suite de l'éclosion de SRAS, l'importance de l'utilisation des médias d'actualité en complément des systèmes nationaux de surveillance de santé publique plus traditionnels a été reconnue par l'OMS et ses États membres (12,13). L'éclosion de SRAS a abouti à des révisions du Règlement sanitaire international (RSI) (14) exigeant que les pays signalent et contrôlent les éclosions potentiellement préoccupantes au niveau international afin de renforcer la sécurité mondiale en matière de santé publique. Le Règlement note que l'OMS peut tenir compte de rapports émanant de sources autres que les notifications et les consultations officielles dans son évaluation d'un événement de santé publique potentiel émergent (14). Après l'éclosion de SRAS, de nombreux pays ont adopté les résultats du RMISP dans le but d'élargir leur capacité de surveillance (4,15).

Au fil des ans, le RMISP a continué à détecter les signaux précoces d'éclosions préoccupantes à l'échelle internationale, comme la grippe pandémique H1N1 en 2009 (16). Les premiers rapports en espagnol à ce sujet faisaient état d'une éclosion d'une maladie respiratoire inhabituelle dans l'État de Veracruz, au Mexique, qui avait provoqué deux décès.

En avril 2012, le RMISP a repéré huit cas d'une maladie respiratoire inconnue et un décès en Jordanie. Le RMISP a envoyé une alerte pour informer ses partenaires, incluant l'OMS, de ces cas. Suite à une enquête plus poussée et les résultats d'une analyse rétrospective de laboratoire, une éclosion du coronavirus du syndrome respiratoire du Moyen-Orient (désormais appelé CoV-SRMO) fut confirmée. Un avis du RSI fut publié en novembre 2012. Le RMISP s'est vu attribuer le mérite d'avoir été le premier à envoyer une alerte concernant ce nouveau virus émergent.

### Surveillance en continu

Le RMISP s'est avéré utile aussi bien pour la détection précoce que pour la surveillance en continu. La surveillance en continu des événements est cruciale pour bien connaître la situation, en ce qui concerne l'évolution d'une éclosion et les stratégies d'intervention et d'atténuation mises en œuvre par les communautés locales, nationales et internationales. Parmi les exemples de connaissance de la situation en matière de stratégies d'atténuation, on peut citer la capacité du RMISP à repérer les annulations de vols ou de croisières, les nouveaux avis aux voyageurs, les procédures de dépistage sanitaire aux frontières ou les interdictions commerciales. Ce processus a été beaucoup plus efficace que de communiquer individuellement avec les sociétés de transport commercial, les agences de voyage et les aéroports.

Par exemple, pendant la grippe pandémique H1N1, le RMISP a été utilisé comme source de renseignements par l'Organisation mondiale du commerce afin de surveiller l'ampleur et l'effet des interdictions commerciales (17). De même, durant l'intervention récente contre l'éclosion du virus Ebola en Afrique occidentale, le RMISP a permis de connaître la situation sur le plan des annulations de vol, des avis aux voyageurs et des procédures de dépistage sanitaire aux frontières.

## Prochaines étapes

### Nouvelles sources de données potentielles

Internet, le courrier électronique, les téléphones intelligents et les médias sociaux se sont développés rapidement depuis la création du RMISP. Par conséquent, de nouvelles sources potentielles de données massives ont émergé qui peuvent être analysées pour détecter les signaux précoces d'éclosions de maladies infectieuses. Les médias sociaux (tels que Twitter et Facebook) ont enregistré une croissance exponentielle au cours des dix dernières années, et ces plateformes créent d'énormes quantités de contenu et de données générés par les utilisateurs (18).

Ces divers médias sociaux constituent de nouvelles sources de données potentielles pour le RMISP. De plus, d'autres organisations ont commencé à exploiter les ressources des médias sociaux pour améliorer la surveillance des maladies (18). Par exemple, l'application Suivi de la grippe de Google surveille les comportements de recherche en ligne pour détecter des signes précurseurs de la grippe (19), des chercheurs ont utilisé Facebook pour contribuer à prévoir les résultats en matière de santé au niveau de la population locale (20), Twitter a été utilisé comme une vaste source de données pour surveiller les tendances de la santé pendant une éclosion de grippe aviaire (21), et les téléphones cellulaires ont servi à mesurer les schémas de mobilité humaine dans le contexte de la transmission du paludisme dans les pays en développement (22).

Les médias sociaux ont contribué à l'amélioration des interventions d'urgence en permettant de saisir des données en temps réel sur la santé des communautés (23) et la réaction du public à un événement (24). Par exemple, l'utilisation de téléphones intelligents et de Twitter au Nigeria pendant l'éclosion d'Ebola en Afrique occidentale a aidé à repérer une éclosion dans une nouvelle région trois jours avant qu'elle soit annoncée par l'OMS (25).

Au nombre des autres nouvelles applications figurent les systèmes d'externalisation à grande échelle, qui saisissent les symptômes soumis de façon volontaire par le grand public à l'aide d'Internet ou des réseaux de téléphone cellulaire et qui les compilent rapidement et fournissent un retour sur ces données en temps quasi réel. Ces systèmes ont été utilisés par des applications participatives de surveillance des maladies infectieuses telles que Flu Near You (26) et DoctorMe (27).

Cependant, l'utilisation des médias sociaux comme sources de données pose quelques problèmes inhérents. L'une des principales difficultés associées aux données massives en général et au contenu des médias sociaux en particulier est le « rapport signal/bruit » qui peut augmenter considérablement le nombre potentiel de faux positifs et de faux négatifs. Avec l'abondance de discussions et de gazouillis autour de l'éclosion du virus Ebola en Afrique occidentale, par exemple, il était difficile de faire la distinction entre les véritables signaux préoccupants et la pléthore de messages à laquelle on pourrait autrement s'attendre pendant un tel événement. En outre, certains médias sociaux, notamment les gazouillis qui sont limités à 140 caractères, peuvent ne pas fournir assez d'information contextuelle pour permettre de repérer un signal fiable (28).

Une autre difficulté associée à l'utilisation des médias sociaux est la représentativité. Tout le monde n'a pas accès à un téléphone intelligent et, par conséquent, les données des plateformes de médias sociaux ne peuvent refléter que la partie de la population qui en utilise un (28). La technologie mobile se répand considérablement, ce qui pourrait contribuer à résoudre les préoccupations relatives à la représentativité (29).

Enfin, l'utilisation des médias sociaux soulève des considérations éthiques associées aux droits des individus, notamment des problèmes liés à la protection des renseignements personnels (2).

### Amélioration de l'analyse des données

Le RMISP pourrait non seulement élargir ses sources de données, mais également faire progresser ses capacités d'analyse des données. Des méthodes avancées de calcul et de vérification pour améliorer la sensibilité et la spécificité des signaux détectés sont actuellement envisagées (30). La question de savoir si

un meilleur traitement des données pourrait réduire la nécessité d'avoir recours à une équipe multilingue et pluridisciplinaire est également à l'étude. Le RMISP évalue et perfectionne en permanence les agrégateurs et les algorithmes utilisés, ce qui pourrait potentiellement aboutir à des formes plus avancées d'intelligence artificielle. La poursuite de l'amélioration de la capacité analytique du RMISP permettra la gestion, l'intégration, l'analyse et l'interprétation solides de volumes de données de plus en plus importants et complexes (31).

## Conclusion

Le Réseau mondial d'information en santé publique du Canada a été parmi les premiers à adopter les données massives et, en tant que ressource mondiale en continu, il aide les pays à satisfaire les exigences en matière de capacité de surveillance des événements pour la détection précoce et le signalement des éclosions de maladies infectieuses et des autres événements préoccupants à l'échelle internationale. Les progrès continus des données massives, notamment l'utilisation des médias sociaux et des téléphones intelligents, ainsi que les améliorations de la capacité d'analyse constituent des occasions de renforcer encore davantage le RMISP. Dans l'ensemble, les approches fondées sur les données massives sont devenues une composante essentielle des efforts en matière de santé publique à l'échelle locale, nationale et internationale pour détecter, signaler et contrôler les éclosions émergentes.

## Remerciements

Nous souhaitons remercier toute l'équipe du RMISP. Le Réseau doit son succès à leur travail assidu, à leur dévouement et à leur constante collaboration. Leur soutien et leurs conseils ont été grandement appréciés et ont fortement contribué à la rédaction de cet article.

## Conflit d'intérêts

Aucun.

## Financement

Le RMISP est financé par l'Agence de la santé publique du Canada.

## Références

(1) George G, Haas MR, Pentland A. Big Data and management. Acad Manag J. 2014;57(2):321-6.
(2) Vayena E, Salathé M, Madoff LC, Brownstein JS, Bourne PE. Ethical challenges of Big Data in public health. PLoS Comput Biol. 2015;11(2):e1003904.
(3) Muchaal P, Meganath K, Landry L, Aramini J. Evaluation of a national pharmacy-based syndromic surveillance system. Can Comm Dis Rep. 2015 41;9:204-210.
(4) Keller M, Blench M, Tolentino H, Freifeld CC, Mandl KD, Mawudeku A, et al. Use of unstructured event-based reports for global infectious disease surveillance. Emerg Infect Dis. 2009 May;15(5):689-95.
(5) World Health Organization. [Internet] Epidemic intelligence - Systematic event detection. Geneva: World Health Organization; 2015. Available from: http://www.who.int/csr/alertresponse/epidemicintelligence/en/.
(6) Mykhalovskiy E, Weir L. The Global Public Health Intelligence Network and early warning outbreak detection: A Canadian contribution to global public health. Can J Public Health. 2006 Jan-Feb;97(1):42-4.
(7) McAfee A, Brynjolfsson E, Davenport TH, Patil D, Barton D. Big Data. The management revolution. Harvard Bus Rev. 2012 Oct;90(10):61-7.
(8) Hay SI, George DB, Moyes CL, Brownstein JS. Big Data opportunities for global infectious disease surveillance. PLoS Med. 2013;10(4):e1001413.
(9) Mawudeku A, Blench M, Boily L, St John R, Andraghetti R, Ruben M. The Global Public Health Intelligence Network. In: Infectious Disease Surveillance, Second Edition. New York: John Wiley and Sons; 2013. pp. 457-69.
(10) Heymann DL, Rodier G. Global surveillance, national surveillance and SARS. Emerg Infect Dis. 2004;10(2):173-5.

(11) Mawudeku A, Blench M. Global Public Health Intelligence Network (GPHIN) In: Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: August 2006. Cambridge, MA, USA.
(12) Davies SE. Nowhere to hide: Informal disease surveillance networks tracing state behaviour. Global Change, Peace & Security. 2012;24(1):95-107.
(13) Davies SE, Youde JR. The Politics of surveillance and response to disease outbreaks: The new frontier for states and non-state actors. Burlington VT: Ashgate Publishing, Ltd.; 2015.
(14) World Health Organization. International Health Regulations. 2008. Second Edition. Geneva: WHO; 2008.
(15) Baker MG, Fidler DP. Global public health surveillance under new international health regulations. Emerg Infect Dis. 2006 Jul;12(7):1058-65.
(16) Warren AP, Bell M, Budd L. Surveillance networks and spaces of governance: Technological openness and international cooperation during the 2009 H1N1 pandemic. Washington: Association of American Geographers; 2010.
(17) Lamy P. Report to the TPRB from the Director-General on the financial and economic crisis and trade-related development. Geneva: World Trade Organization; 2009.
(18) Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res. 2013 Jul 18;15(7):e147.
(19) Davidson MW, Haim DA, Radin JM. Using networks to combine Big Data and traditional surveillance to improve influenza predictions. Sci Rep. 2015;5:8154.
(20) Gittelman S, Lange V, Gotway Crawford CA, Okoro CA, Lieb E, Dhingra SS, et al. A new source of data for public health surveillance: Facebook likes. J Med Internet Res. 2015 Apr 20;17(4):e98.
(21) Fung IC, Wong K. Efficient use of social media during the avian influenza A (H7N9) emergency response. Western Pac Surveil Response J. 2013;4(4):1.
(22) Buckee CO, Wesolowski A, Eagle NN, Hansen E, Snow RW. Mobile phones and malaria: Modeling human and parasite travel. Travel Med Infect Dis. 2013;11(1):15-22.
(23) Merchant R, Elmer S, Lurie, N. Integrating social media into emergency-preparedness efforts. N Engl J Med. 2011 July 28;365(4):289-91.
(24) Fung IC, Fu K, Ying Y, Schaible B, Hao Y, Chan C, et al. Chinese social media reaction to the MERS-CoV and avian influenza A (H7N9) outbreaks. Infect Dis Poverty. 2013;2(1):1-12.
(25) Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? Am J Infect Control. 2015;43(6):563-71.
(26) Wojcik OP, Brownstein JS, Chunara R, Johansson MA. Public health for the people: Participatory infectious disease surveillance in the digital age. Emerg Themes Epidemiol. 2014 Jun20;11:7,7622-11-7. eCollection 2014.
(27) Susumpow P, Pansuwan P, Sajda N, Crawley AW. Participatory disease detection through digital volunteerism: How the DoctorMe application aims to capture data for faster disease detection in Thailand. Proceedings of the companion publication of the 23rd International Conference on World Wide Web Companion; International World Wide Web Conferences Steering Committee; 2014.
(28) Guy S, Ratzki-Leewing A, Bahati R, Gwadry-Sridhar F. Social media: A systematic review to understand the evidence and application in infodemiology. In: Electronic Healthcare. New York: Springer; 2012. p. 1-8.
(29) e Geneva: International Telecommunications Union; 2013. Available from: http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013-e.pdf.
(30) Kostkova P. A roadmap to integrated digital public health surveillance: The vision and the challenges. Proceedings of the 22nd International Conference on World Wide Web Companion; International World Wide Web Conferences Steering Committee; 2013.
(31) Velasco E, Agheneza T, Denecke K, Kirchner G, Eckmanns T. Social media and internet-based data in global systems for public health surveillance: A systematic review. Milbank Q. 2014;92(1):7-33.

Commentaire

# Les données massives font évoluer la lutte contre les maladies infectieuses

**Links MG**[1, 2*]

[1]Centre de recherches de Saskatoon, Agriculture et Agroalimentaire Canada, Saskatoon (Saskatchewan)
[2]Département des sciences informatiques, Université de la Saskatchewan, Saskatoon (Saskatchewan)

*Correspondance : Matthew.Links@usask.ca

## Résumé

**Les données massives étaient traditionnellement associées aux passionnés d'informatique et aux entreprises commerciales, mais elles sont devenues bien ancrées dans de nombreuses disciplines scientifiques, notamment la prévention et le contrôle des maladies infectieuses. L'utilisation des données massives a permis de déterminer les tendances des maladies et de retracer, voire de prédire, les origines des éclosions. Le volume des données massives ne diminue pas. Nous sommes confrontés à des difficultés qui nous poussent à perfectionner notre capacité d'analyse pour faire face à l'énorme « rapport signal/bruit » au moyen d'une puissance de calcul adéquate et d'équipes pluridisciplinaires capables de gérer des quantités de données en constante augmentation. Grâce aux données massives, il sera également possible de créer de futures applications de traitement *sur mesure* (ou personnalisé).**

## Introduction

Les données massives semblent être une évolution récente qui, pour beaucoup, est liée à des expressions telles que l'« informatique en nuage ». Or, le terme est né avant le changement de millénaire, dans les années 1990, lorsque John Mashey était informaticien en chef chez Silicon Graphics (SGI). À l'époque, SGI était à l'avant-garde de l'infographie et luttait pour faire face à des besoins de calcul en considérable expansion qui dépassaient les capacités du matériel disponible. Mashey a rédigé à la fin des années 1990 un exposé qui annonçait la collision imminente entre les données massives et la performance de calcul (1).

Aujourd'hui, l'expression « données massives » est couramment utilisée pour décrire des situations dans lesquelles les volumes de données sont caractérisés par des propriétés incluant, entre autres, la taille, la vitesse d'évolution dans le temps et la nature hétérogène des données elles-mêmes (2). Le terme « données massives » fait généralement référence à d'importants volumes de données qui peuvent être structurés (p. ex. bases de données relationnelles) ou non (p. ex. flux d'actualités Twitter) et qui sont analysés pour en extraire de l'information. Bien que, par le passé, les données massives soient nées dans les domaines des sciences informatiques, de la statistique et de l'économie (3), elles ont depuis été plus largement adoptées dans toutes les disciplines scientifiques.

On dit généralement qu'un problème implique des données massives lorsque le volume est si important qu'il devient impossible de convertir les données en connaissances. Dans le cas de la recherche sur les maladies infectieuses, les données massives ont des répercussions considérables. La capacité à effectuer un suivi des maladies en temps réel et à prévoir les éclosions s'est fondée sur des données non structurées pour faire évoluer la manière dont les maladies infectieuses sont prises en charge. Par exemple, grâce à l'utilisation de diverses sources d'actualité, le RMISP a été utilisé pour signaler de façon précoce de nouvelles infections (telles que le SRAS et le CoV-SRMO) qui a servi de base à une intervention de santé publique en réponse aux éclosions qui ont suivi (4).

Les données structurées sont particulièrement utiles lorsqu'il s'agit de compiler des renseignements provenant de multiples sources en fonction d'une structure prévisible des données. Dans le cas de la surveillance de santé publique, il est désormais possible de chercher des données structurées pouvant

servir de source d'information en remplacement d'une confirmation en laboratoire ou d'une déclaration de cas émise par un médecin. Muchaal et ses collègues ont démontré que l'utilisation de médicaments constitue une source possible d'information précoce de substitution (5).

### Quelle est l'ampleur des données massives?

Il est difficile de se rendre compte d'à quel point le volume de données massives est réellement important. Dans un récent article, Stephens et ses collègues ont replacé les données massives de deux disciplines dans un contexte relatif (6). Bien que la palme du volume faramineux de données revienne actuellement aux études d'astronomie, Stephens et ses collègues ont indiqué que la génomique devrait atteindre le volume des données en astronomie d'ici à 2025. Pour avoir une idée de l'échelle, les données de la génomique en 2025 pourraient être équivalentes à 8 milliards de fois la capacité des plus gros iPhones disponibles aujourd'hui (128 Go de mémoire en 2015), soit l'équivalent des données contenues dans un iPhone par personne sur Terre, chaque année.

## Les données massives et les origines d'une éclosion

L'application de la génomique aux maladies infectieuses peut contribuer à déterminer d'où proviennent réellement les éclosions de maladies infectieuses. Le travail d'enquête réalisé pour faire face à une éclosion de rougeole pendant les Jeux olympiques de 2010 en constitue un bon exemple (7). Au moyen du séquençage complet du génome, Gardy et ses collègues ont réussi à identifier exactement bon nombre des cas signalés (30 sur 82). Un constat important a été que l'éclosion était causée par plusieurs types de virus de la rougeole. Alors que le génotypage traditionnel du virus de la rougeole se concentre sur les séquences de la phosphoprotéine et de l'hémagglutinine, deux gènes spécifiques utilisés pour distinguer les isolats, Gardy et ses collègues ont démontré que d'autres gènes de la rougeole contenaient des variations supplémentaires qui pouvaient être utilisées pour définir plus précisément les lignées virales.

## À venir

L'une des dernières applications des données massives consiste en ce que Jennifer Gardy a appelé le traitement *sur mesure* (ou personnalisé) (8). Par exemple, les approches examinant le génome complet appliquées à des milliers d'isolats peuvent permettre de repérer une variation génomique associée à des phénotypes antimicrobiens de *Mycobacterium tuberculosis* (9). À l'avenir, cette méthode pourrait être appliquée de manière généralisée pour déterminer les meilleurs traitements à adopter contre les bactéries présentant une résistance aux antimicrobiens.

### Comment allons-nous interpréter toutes ces données?

En dépit de tout ce potentiel de progrès, des problèmes clés restent à surmonter pour l'utilisation des données massives dans toutes les disciplines. À mesure que les données deviennent de plus en plus massives, il devient plus difficile de les interpréter : soit l'intégration des données est tellement complexe qu'elle est difficile à suivre sans de sérieuses ressources de calcul, soit leur simple échelle dépasse l'entendement.

Avec l'utilisation du texte non structuré provenant des flux d'actualité analysés aux fins de connaissance et de surveillance des maladies, les données massives servent à trouver du sens dans un déluge de bruit. L'échelle de la quantité de texte analysée, entre les articles scientifiques publiés dans les revues, les nouvelles décrivant de nouveaux enjeux et les gazouillis de 140 caractères, est vraiment intimidante, si l'on considère qu'un demi-milliard de micromessages sont publiés sur Twitter par jour (6). Pour compliquer encore davantage les choses, la surveillance des maladies ne se limite pas simplement à compiler des flux d'actualité. Au contraire, la clé d'une stratégie significative de surveillance des maladies axée sur les données massives est la détermination du risque potentiel. Ainsi, les approches telles que le RMISP (4) sont cruciales pour que les pays et la communauté internationale soient prêts à intervenir en cas d'éclosion de maladie. Comme les données massives continuent à croître à une vitesse exponentielle, tenter d'améliorer notre capacité à les analyser devient une quête du Saint-Graal en constante évolution.

Trop souvent, les données massives sont acquises dans le cadre d'une étude pluridisciplinaire et remises à une seule personne (p. ex. étudiant diplômé, étudiant postdoctoral ou boursier) en espérant que celle-ci pourra, à elle seule, réussir à comprendre ce que tout cela signifie. Confier à une seule personne la responsabilité de relations de plus en plus complexes émanant de volumes écrasants de données ne constitue tout simplement pas une stratégie viable. C'est pourquoi la tendance est à l'élaboration d'approches pluridisciplinaires pour l'interprétation des données massives (4).

## Conclusion

À moins d'imaginer un âge des ténèbres de l'informatique, il est difficile de croire que le volume des données massives va diminuer. Par conséquent, les disciplines scientifiques ont développé leur capacité à exploiter des volumes de données en constante augmentation. La prudence est de rigueur pour ne pas se faire submerger par les données massives. De fait, c'est l'évolution vers une analyse pluridisciplinaire des données massives qui permet aux équipes de suivre les tendances des maladies et même de prédire les éclosions avant qu'elles n'aient lieu. Les données massives sont en position de se déplacer progressivement de la santé publique au milieu clinique; le traitement *sur mesure* (ou personnalisé) des maladies infectieuses pourrait bientôt être à notre porte.

## Financement

Le programme de recherche du Dr Links a été ou est actuellement financé par Agriculture et Agroalimentaire Canada (AAC), l'Initiative de recherche et développement en génomique (IRDG) du gouvernement du Canada, les Instituts de recherche en santé du Canada (IRSC), le Conseil national de recherches du Canada (CNRC), l'Agriculture Development Fund (ADF) de la Saskatchewan et le Programme canadien pour la sûreté et la sécurité (PCSS) du gouvernement du Canada (auparavant appelé Initiative de recherche et de technologie chimique, biologique, radiologique et nucléaire [IRTC]).

## Conflit d'intérêts

Aucun.

## Références

(1) Mashey J, ed. Big Data and the next wave of infrastress. USENIX Annual Technical Conference; 1998; Monterey, California, USA: Usenix.
(2) Asokan GV, Asokan V. Leveraging "big data" to enhance the effectiveness of "one health" in an era of health informatics. J Epidemiol Glob Health. 2015 Mar 5.
(3) Diebold FX. A personal perspective on the origin(s) and development of 'Big Data': The phenomenon, the term, and and the discipline. Department of Economics, University of Pennsylvania. PIER Working Paper No. 13-003. November 2012.
(4) Dion M, AbdelMalik P, Mawudeku A. Big Data and the Global Public Health Intelligence Network (GPHIN). Can Commun Dis Rep. 2015;41:211-216.
(5) Muchaal PK, Parker S, Meganath K, Landry L, Aramini J. Evaluation of a national pharmacy-based syndromic surveillance system Can Commun Dis Rep. 2015;41:204-211.
(6) Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, et al. Big Data: Astronomical or genomical? PLoS Biol. 2015 Jul;13(7):e1002195.
(7) Gardy JL, Naus M, Amlani A, Chung W, Kim H, Tan M, et al. Whole-genome sequencing of measles virus genotypes H1 and D8 during outbreaks of infection following the 2010 Olympic Winter Games reveals viral transmission routes. J Infect Dis. 2015 Jul 6.
(8) Gardy JL. Towards genomic prediction of drug resistance in tuberculosis. Lancet Infect Dis. 2015 Jun 23.
(9) Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, et al. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: A retrospective cohort study. Lancet Infect Dis. 2015 Jun 23.

## Actualités sur les maladies infectieuses : Les données massives et la prédiction, la prévention et le contrôle des éclosions

Christaki E. **New technologies in predicting, preventing and controlling emerging infectious diseases.** Virulence. Juin 2015, 11:1-8. (*Résumé*)

La surveillance des maladies infectieuses émergentes est essentielle pour repérer rapidement les menaces en matière de santé publique. L'apparition de nouvelles infections est liée à des facteurs humains tels que la densité de population, les voyages et le commerce, ainsi qu'à des facteurs écologiques comme les changements climatiques et les pratiques agricoles. Une multitude de nouvelles technologies devient de plus en plus accessible pour l'identification moléculaire rapide des pathogènes, mais aussi pour le suivi plus précis de l'activité des maladies infectieuses. Les outils de surveillance sur Internet et les méthodes d'information sur les épidémies, utilisés par toutes les grandes institutions de santé publique, ont pour but de faciliter l'évaluation des risques et la détection rapide des éclosions. Cet examen présente les nouvelles méthodes de surveillance des maladies infectieuses à l'échelle régionale et mondiale ainsi que les progrès réalisés en matière de modélisation des épidémies visant à prédire et prévenir les futures menaces liées à des maladies infectieuses.

Semenza JC. **Prototype early warning systems for vector-borne diseases in Europe.** Int J Environ Res Public Health. Juin 2015, 2;12(6):6333-51. doi:10.3390/ijerph120606333. (*Résumé*)

La mondialisation et les changements environnementaux, les déterminants sociaux et démographiques et la capacité des systèmes de santé sont des facteurs importants de maladies infectieuses, qui peuvent aussi agir comme précurseurs des épidémies. C'est pourquoi la surveillance des évolutions de ces facteurs peut contribuer à anticiper, ou même à prévoir, une flambée des maladies infectieuses. Le Réseau européen Environnement et Épidémiologie (E3) a été conçu dans ce but et appliqué à trois études de cas de détection précoce : 1) la pertinence environnementale de la transmission du paludisme en Grèce a été cartographiée afin de cibler les activités de surveillance épidémiologique et entomologique et de contrôle des vecteurs. La transmission du paludisme dans ces régions a cessé en 2013 grâce à ces activités intégrées de préparation et d'intervention, 2) depuis 2010, des éclosions récurrentes de fièvre du Nil occidental sont apparues en Europe du Sud et de l'Est. Il a été prouvé que l'éclosion de 2010 était associée à des écarts de température par rapport à une moyenne sur trente ans. Les facteurs des éclosions suivantes, qui ont été calculés au moyen de modèles de régression logistique à plusieurs variables, comprenaient des anomalies des températures mensuelles en juillet et un indice normalisé d'eau, 3) la dengue est une maladie tropicale, mais une transmission durable est récemment apparue à Madère. Une transmission autochtone a également été observée à plusieurs reprises en France et en Croatie, principalement du fait de l'importation du virus lors de voyages. Le risque d'importation de la dengue en Europe en 2010 a été calculé en fonction du volume de voyageurs internationaux en provenance des régions touchées par la dengue dans le monde entier. Ces prototypes de systèmes de détection précoce indiquent que la surveillance des facteurs des maladies infectieuses peut aider à prévoir les menaces liées aux maladies à transmission vectorielle.

Hay SI, George DB, Moyes CL, Brownstein JS. **Big Data opportunities for global infectious disease surveillance.** PLoS Med. 2013;10(4):e1001413. doi:10.1371/journal.pmed.1001413. (*Résumé*)

Il existe des systèmes permettant de fournir, d'une part, des cartes statiques spatialement continues du risque de maladie infectieuse et, d'autre part, des rapports mis à jour en permanence sur l'occurrence des maladies infectieuses, mais, à ce jour, ces deux fonctions n'ont jamais été combinées. Les nouvelles sources de données en ligne, telles que les médias sociaux, associées à des renseignements environnementaux pertinents sur le plan épidémiologique constituent de nouvelles sources de données précieuses qui peuvent contribuer à la mise à jour « en temps réel » des cartes spatiales. Grâce aux progrès de l'apprentissage machine et à l'utilisation de l'externalisation à grande échelle, il devient possible de concevoir un atlas des maladies infectieuses constamment actualisé. Le libre accès à des cartes dynamiques du risque de maladies infectieuses serait extrêmement utile pour un large spectre de professionnels de la santé, des responsables des politiques chargés de répartir des ressources limitées par ordre de priorité aux cliniciens individuels.

## Actualités sur les maladies infectieuses : Données massives et éthique

Ploug T, Holm S. **Meta consent: A flexible and autonomous way of obtaining informed consent for secondary research**. BMJ. 7 mai 2015;350:h2146. doi:10.1136/bmj.h2146. (*Résumé*)

La croissance rapide de la capacité de stockage, de mise en relation et d'analyse des données sur la santé a ouvert de nouvelles opportunités en matière de recherche. Toutefois, elle soulève également de nouvelles préoccupations sur le plan de l'éthique et de la réglementation. Parmi celles-ci se pose notamment la question centrale des conditions dans lesquelles les travaux de recherche secondaire peuvent utiliser les données qui ont été recueillies dans le cadre des pratiques de base de soins de santé ou pour un projet de recherche particulier. L'utilisation secondaire des données nécessite-t-elle un renouvellement du consentement éclairé des participants initiaux? Jusqu'à présent, il existait trois types de consentement : le consentement dynamique (l'information sur l'utilisation secondaire particulière de données sur la santé ou de tissus est demandée à chaque fois à chaque personne par le biais d'une plateforme sur Internet), le consentement large (le sujet donne son consentement à de futures recherches d'un type particulier en plus du projet de recherche spécifique en cours) et le consentement général (les données pourraient être utilisées sans autre consentement). Nous proposons un métaconsentement, ce qui signifie que les personnes pourront choisir comment elles souhaitent accorder leur consentement à de futures recherches secondaires sur les données recueillies par le passé ou sur les données qui seront stockées à l'avenir; le métaconsentement est donc à la fois rétrospectif et prospectif. Le métaconsentement est une procédure de consentement véritablement individuelle, qui tient compte des différences au sein de la population sur le plan de l'intérêt personnel et du degré de confiance dans les chercheurs. Le risque de routinisation est réduit, car les personnes peuvent limiter les demandes qu'elles reçoivent uniquement aux catégories de recherche qui les intéressent vraiment. Sa mise en œuvre en ligne rend le métaconsentement facile à révoquer ou à modifier.

Mittelstadt BD, Floridi L. **The ethics of Big Data: Current and foreseeable issues in biomedical contexts**. Sci Eng Ethics. 23 mai 2015. [Publication électronique et non imprimée] (*Résumé*)

La capacité de collecte et d'analyse de données augmente de façon exponentielle. Désignée sous le nom de « données massives », cette tendance scientifique, sociale et technologique a contribué à créer des quantités déstabilisantes d'information, qui peuvent remettre en cause les normes sociales et éthiques acceptées. Comme c'est souvent le cas avec les progrès scientifiques et technologiques de pointe, la compréhension des répercussions éthiques des données massives est à la traîne. Au moyen d'une méta-analyse de la littérature, un exposé thématique est présenté pour aider les éthiciens, les scientifiques des données, les organismes de réglementation et les autres intervenants à s'y retrouver dans ce que l'on sait ou suppose déjà sur les risques éthiques liés à ce phénomène émergent et innovant. Cinq principaux aspects préoccupants sont décrits : 1) le consentement éclairé, 2) la protection des renseignements personnels (y compris l'anonymisation et la protection des données), 3) la propriété des données, 4) l'épistémologie et l'objectivité, et 5) les « fractures des données massives » qui se créent entre ceux qui disposent des ressources nécessaires pour analyser des ensembles de données de plus en plus vastes et ceux qui ne les ont pas. Les auteurs évoquent ensuite six autres domaines préoccupants qui, bien qu'ils soient liés à ce phénomène, n'ont pas encore suscité de débat approfondi dans la littérature existante : 6) les dangers d'ignorer les préjudices éthiques au niveau collectif, 7) l'importance de l'épistémologie dans l'évaluation de l'éthique des données massives, 8) la nature évolutive des relations fiduciaires qui deviennent de plus en plus saturées de données, 9) la nécessité de faire la distinction entre les pratiques « universitaires » et « commerciales » d'utilisation des données massives sur le plan des préjudices potentiels pour les personnes concernées, 10) les futurs problèmes liés aux droits de propriété intellectuelle générés par l'analyse d'ensembles de données agrégées, et 11) la difficulté à accorder des droits d'accès significatifs aux personnes individuelles concernées qui ne disposent pas des ressources nécessaires. Ensemble, ces onze thèmes constituent un cadre critique complet pour guider l'évaluation et la gouvernance éthiques des pratiques émergentes en matière de données massives.